DOCUMENT RESUME

ED 034 594 PS 002 700

| | |
|---|---|
| AUTHOR | Rankin, Richard J.; Henderson, Ronald W. |
| TITLE | Standardized Tests and the Disadvantaged. |
| INSTITUTION | Arizona Univ., Tucson. Arizona Center for Early Childhood Education.; National Lab. on Early Childhood Education. |
| SPONS AGENCY | Office of Education (DHEW), Washington, D.C. Div. of Educational Labs. |
| REPORT NO | NLECE-70706-A-CD-U-13 |
| PUB DATE | Nov 69 |
| CONTRACT | OEC-3-7-70706-3118 |
| NOTE | 13p. |
| EDRS PRICE | EDRS Price MF-$0.25 HC-$0.75 |
| DESCRIPTORS | Cultural Factors, *Culturally Disadvantaged, *Intelligence Tests, Mexican Americans, Predictive Ability (Testing), Preschool Children, Sex Differences, *Test Reliability |

ABSTRACT

The purpose of this paper is to evaluate the reliability of the Wechsler Preschool and Primary Scale of Intelligence and to measure whether this reliability is affected when subjects are from a disadvantaged group. The subjects were 25 male and 24 female 5 1/2-year-old poor Mexican-Americans. Generally, the Wechsler Preschool Scale showed high reliability with this sample, full scale reliability being .95. The impact of disadvantage can be seen by comparing the scale scores of this group with the results reported for an Anglo-American standardization group. The Mexican-American group falls below the general mean in all subtests, noticeably in the verbal section and most notably in the information and similarities sections of the Wechsler Scale. The high reliability of the scale suggests implications for testing children from ethnic minorities. Since most of the children had limited facility with English, study results will encourage researchers who want an accurate measure of intellectual skills required for successful performance in technical cultures. Measurement within a disadvantaged group may not require new tests to predict skills but may depend upon the use of tests sampling existing known factors and utilizing them to predict within groups. Norms should be established for the specific group tested. (JF)

ED034594

Document 70706-A-CD-U-13

Printed November, 1969

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

# Standardized Tests and the Disadvantaged

Richard J. Rankin, Ronald W. Henderson

(Arizona)

The research or work reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare through the Arizona Center for Early Childhood Education, a component of the National Laboratory on Early Childhood Education, contract OEC-3-7-70706-3118.

Contractors undertaking such work under Government sponsorship are encouraged to express freely their professional judgment in the conduct of work. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

PS 002700

# STANDARDIZED TESTS AND THE DISADVANTAGED

Richard J. Rankin, Ronald W. Henderson

The general purpose of this paper is to evaluate the reliability of a new individual intelligence scale, the Wechsler Preschool and Primary Scale of Intelligence. A secondary purpose is to consider the idea that perhaps carefully administered and specifically renormed conventional intelligence tests might be a better solution to cultural-bias problems than the culture-free intelligence test. In one sense, this paper presents an independent evaluation of the reliability of the Wechsler Preschool Scale with this evaluation combined with measurement of the impact of disadvantaged group membership upon the tests reliability.

The Wechsler Pre-Primary Scale is basically a downward extension and modification of the Wechsler Intelligence Scale for Children. The most noticeable distinction between the two tests is the elimination of the digit span subtest and the replacement of the coding subtest with a modified form of a coding instrument called animal house. The Wechsler Preschool Scale was designed to correct long apparent differences in the lower age ranges of the WISC while extending this lower range of the test from the WISC lower limit of five years to a lower limit of four years. The upper age range of the Wechsler Preschool Scale is six and one-half.

There are two major deficiencies apparent with the WISC which are so serious that the Psychological Corporation strongly recommends using

the Wechsler Preschool Scale in place of the WISC, at those ages where the two scales overlap. The first major deficiency with the WISC is that it has too high a floor with younger children. Because of its origins in the original Wechsler Bellevue, there are really no easy items at the lower end, and it has often been pointed out that paradoxically a child can obtain scale score even though he is incapable of performing correctly even one question on certain subtests. For example, a scale score of six can be obtained by a five-year-old child who gets no arithmetic problems right. Thus undoubtedly leads to low reliability with duller subjects at the younger age levels. Even though a WISC has norms for five-year-olds, reliabilities are reported for subjects no younger than seven and one-half, at which level the reliability reported is .88 for verbal and .86 for performance. These levels are below the minimum recommended by many psychologists for individual diagnosis.

The second major difficulty of the WISC shared with the Stanford Binet is that the scale was standardized on all white population and, thus, may not be suitable for use with minority groups because they are not represented in the norming sample. The Wechsler Preschool attempted to overcome these difficulties by standardizing the scale upon a sample of the population proportionately representative of the population of the United States in terms of ethnic background and socioeconomic level. The problem of

reliability with the WPPSI seems to be fairly well solved as the manual reports reliabilities coefficients for the instrument, varying from a low of .96 to a high of .97 at all six age levels. The effectiveness of the scale as a prediction instrument remains to be seen.

A major problem in intellectual measurement has always been speculation concerning the impact of item biasing upon the validity of the obtained measurement. This problem is aggravated when the investigator takes measured intelligence to be somehow related to genetic endowment rather than recognizing that it is simply a statistical abstraction sometimes useful for predicting specific achievements.

The problem of item biasing was first pointed out by Binet in about 1905 and has reoccurred frequently in the literature. Multitudes of studies have shown that various minority groups achieve scores deviating from the mean upon intelligence tests, and this indication of bias has been pointed out as a reason to suspect the efficiency of tests as a predictor of school success. A question needing further evaluation and the one that has been much ignored is the impact of minority group membership on the reliability of an instrument when that instrument was standardized upon a sample in which the minority group is represented. Investigation of the WPPSI may solve some of these problems.

The subjects are 25 male and 24 female Mexican-American children from neighborhoods which were designated as poverty areas in Tucson,

Arizona. This sample was composed of all the five-and-one-half-year-olds, plus or minus six months, contained in a larger group of children, both Mexican-American and Anglo-American, who had been tested with a Wechsler Preschool Scale as a pretreatment measure composed for use as an evaluation device in an experimental intervention program. The range of chronological ages was restricted in order to reduce the possibility of obtaining spuriously high reliabilities which are generated by correlating split-half raw scores in a population of high variance in chronological age. Further, the age restriction makes possible comparison of the Mexican-American sample with an age range specified in the reliability section of the Wechsler Preschool Manual. The range five and one-half was chosen because this produces the largest subsample out of a population of five, five and one-half, and six year olds. The mean intelligence of the five and one-half year olds sample on the Wechsler Preschool was total I.Q. 80, Verbal 74, and Performance 91. There were no significant differences between males and females in total or part I.Q. scores.

Test Administration and Scoring

Tests were administered by full time research assistants who had been provided with intensive training specific to the Wechsler Preschool instrument over a period of approximately three weeks. During the first week, one of the authors conducted training sessions of eight hours duration each

day. Trainees administered tests under supervision and practice sessions were recorded on video tape to provide each trainee with the opportunity to observe the performance of himself and other trainees. Each practice test was checked by the trainer and group critiques served to further reduce individual variability in a test administration and scoring. Following the week of direct instruction, trainees worked in pairs and they continued to spend their full time administering practice tests. During each administration, one member of the pair served as a monitor and members of the pairs were rotated, so each trainee had an opportunity to observe and be observed by every other trainee.

All tests were double checked, again rotating checkers to eliminate scoring errors. Reoccurring difficulties on judgment items were resolved in training meetings. Administration of tests which were used in the present analysis took place over a period of 39 days immediately following the training and practice periods. With the tests included in the study, a double checking procedure was followed with each test scored by two individuals.

PS002700

Analysis of the Data

The procedures for determining the reliability of the Wechsler Preschool Scale follow as closely as possible those worked out in the test manual. Each subtest was split using an odd-even technique and the part

scores were correlated with the resultant R's corrected using the Spear-man Brown prophecy formula. The animal house or coding test was not included in this investigation as it is speeded and split half formulas are not appropriate. The total and subscales were analyzed separately for males and females, to determine if significant differences in reliabilities obtained between the sexes. Examination of the table indicates that with the total test there is no difference between the reliability of the Wechsler Preschool Scale when determined with the sample of five and one-half year olds reported in the manual. With both samples, the reliability coefficient remains about .95 for the total test score. One subtest does show a reliability coefficient significantly lower than that reported in the manual for females. This is the similarity test. The Mexican-American females also show a lower reliability than the Mexican-American males. There is a considerably lower reliability obtained by Mexican-American females and the norming sample in terms of math. The difference between males and females is not significant, but the difference between Mexican-American females and the correlation reported in the manual is significant at the 5 per cent level. The reliability of the total verbal scale with the Mexican-American sample of females is less than the reliability reported in the manual. This difference is significant and probably reflects the lowered reliabilities of the arithmetic and similarity tests. The practical significance of this difference in verbal scale can be seen by

converting the reliability coefficient of the female population to a standard error of measurement, which is found to be 5.00 I.Q. points compared with 3.40 with the sample reported in the manual, which consists of males and females combined. It would be of interest to know what the sex differences in reliability are for the sample used to standardize the test. Generally, in terms of reliability the Wechsler Preschool Scale is shown to be highly reliable with this sample of disadvantaged Mexican-American youths. The impact of disadvantage on these children can be seen by examining the table of scale scores from a Mexican-American group and comparing the results with a scale score mean of 10 and standard deviation of 3 reported for the standardization group. It can be seen that the Mexican-American group falls below the general mean in all subtests more noticeably in the verbal section and most notably in the information and similarities sections of the Wechsler Preschool Scale. There are no significance difference between performance of the male and female sample in terms of intellectual functioning.

A small sample validation of the Wechsler Preschool was possible comparing it with the picture vocabulary test produced by Ammons and Ammons. The Wechsler Preschool Manual reports correlations with the Peabody Picture Vocabulary Test Form A of .57 with the Verbal and .44 with the Performance sections. Utilizing 18 Mexican-American subjects from the sample who had been given the Wechsler Preschool and the Full

Range Vocabulary Test correlations of .61 and .33 were found for the verbal and performance scales respectively, omitting animal house. With 21 nondisadvantaged Anglo children in a control group, the correlation between the Wechsler Preschool verbal and performance and the Full Range Vocabulary Test were found to be .51 and .09 respectively. Generally, the same level of relationship between the Wechsler Preschool and the Picture All Vocabulary Test is found in three groups.

The high reliabilities indicated for Mexican-American children in these data suggest some implications for the testing of children from ethnic minorities. Over the past few years, many psychologists have stressed the role of experience in the development of skills and concepts reflected in intellectual performance. McCandless suggests that intelligence is subject to the laws of learning and that intelligence can be thought of as the "level of problem solving ability which has been reached by the tested person." Vernon refers to intelligence as "the totality of concepts and skills, the techniques or plans for coping with problems which have crystallized out of the child's previous experience." It seems undeniable that individuals intellectual skills emerge through interaction with the experiences available and valued in his cultural or social milieu. Therefore, we would expect children reared in settings which differ in significant ways from the culture of the norming sample to display below average performance on standardized tests. This says nothing about their genetic inheritance. Does

this necessarily mean that such tests are without value for children from ethnic minorities or from lower socioclass backgrounds. Eells has advocated the development of culture fair tests. Experience has indicated, however, that such tests do not necessarily reflect higher performance for the child from a lower social class or ethnic minority background. Kuhlmann and Ward found the discrepancy in performance of lower class and middle class children to be just as great on the Davis Eells games as on the Kuhlmann Finch. Fowler reported larger correlations for social class in the Davis Eells games than for the conventional tests and social class. Thorndike and Hagen report that tests of this type are relatively unreliable. It might be concluded that even if a culture free test were desirable, adequate ones have yet to be developed. The mention of the desirability of such tests raises an important question. If we had a culture free test, what would it predict? McCandless has suggested that intelligence tests of the conventional type do predict skills that are valued in this culture if the culture free tests existed it would be of limited usefulness. Our present data demonstrates that the Wechsler Pre-Primary Scale carefully administered and scored is a reliable test for the sample of Mexican-American children. The majority of these children had very limited facility with the English language. This should be encouraging to researchers who want an accurate measure of the current status of intellectual skills required for successful performance in technical cultures. A message seems to be

emerging in current research and it might be that measurement within a disadvantaged group may not require new kinds of tests to predict skills we value, but may rather depend upon the use of tests sampling existing known factors and utilizing them to predict within groups. Perhaps more progress in mental measurement will be made when we produce norms specific for the groups we wish to predict for and spend less effort attempting to provide general factor tests with one all-purpose set of norms.

Condensed Data for WPPSI Reliability and Validity Study
Five-and-one-half-year-old Mexican-Americans

| | Male | Female | Total | Manual |
|---|---|---|---|---|
| Information | .81 | .77 | .77 | .81 |
| Vocabulary | .91 | .93 | .91 | .85 |
| Arithmetic | .86 | .67 | .78 | .86 |
| Similarities | .87 | .52 | .77 | .82 |
| Comprehension | .89 | .84 | .87 | .84 |
| Picture Completion | .91 | .94 | .92 | .86 |
| Mazes | .81 | .85 | .85 | .91 |
| Geometric Design | .84 | .89 | .89 | .84 |
| Block Design | .91 | .88 | .89 | .85 |
| Verbal I.Q. | .97 | .90 | .95 | .95 |
| Performance I.Q. | .91 | .93 | .92 | .95 |
| Full Scale | .97 | .91 | .95 | .97 |

n = 49, 25 males; 24 females

Scale Score Averages

| | Verbal | | | | | Performance | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Info | Vocab | Math | Sim | Comp | Picture Comp | Maze | Geo Design | Blocks |
| Male | 5 | 6 | 7 | 5 | 6 | 7 | 9 | 8 | 10 |
| Female | 5 | 6 | 7 | 5 | 7 | 7 | 8 | 10 | 9 |
| Total | 5 | 6 | 7 | 5 | 6 | 7 | 9 | 9 | 9 |

Pick up Scaled Scores from Original and Test Differences.

WPPSI x FRVT inter (s = 18 Mexican-Americans; 21 Anglo)

| | Performance | Verbal |
|---|---|---|
| MA | .33 | .61 |
| AA | .09 | .51 |